Cross-Cultural Cognition Multinational Project-The Second Rosetta Workshop

3 November – 5 November, 2005

Taipei, Taiwan

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
27 SEP 2006	**Conference Proceeding (Technical)**	**29-08-2005 to 31-01-2006**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Cross-Cultural Cognition Multinational Project-The Second Rosetta Workshop	**FA520905P0631**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **National Taiwan University,No.1, Sec. 4, Roosevelt Rd.,Taipei 106, Taiwan,TP,106**	8. PERFORMING ORGANIZATION REPORT NUMBER **CSP-051058**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) **The US Resarch Labolatory, AOARD/AFOSR, Unit 45002, APO, AP, 96337-5002**	10. SPONSOR/MONITOR'S ACRONYM(S) **AOARD/AFOSR**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) **CSP-051058**

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The primary purpose of the workshop was to discuss the findings collected in the first phase of the Rosetta project across four countries: USA, Korea, Japan, and Taiwan (please see Appendix 1 for the agenda of the workshop). A preliminary report of the findings was sent to all members prior to the workshop (see Appendix 2) for the investigators to preview and provide feedback. The second objective was to discuss the future plan for the second phase of the project, with invited observers from Malaysia and India. This report summarizes the proceeding of the workshop and the important suggestions and conclusions related to the results obtained in Phase I

15. SUBJECT TERMS
Cognitive Psychology , Social Psychology, Cultural Psychology, Decision Making

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**7**	

Executive Summary

The primary purpose of the workshop was to discuss the findings collected in the first phase of the Rosetta project across four countries: USA, Korea, Japan, and Taiwan (please see Appendix 1 for the agenda of the workshop). A preliminary report of the findings was sent to all members prior to the workshop (see Appendix 2) for the investigators to preview and provide feedback. The second objective was to discuss the future plan for the second phase of the project, with invited observers from Malaysia and India. This report summarizes the proceeding of the workshop and the important suggestions and conclusions related to the results obtained in Phase I.

3 November

Background

Dr. Helen Klein from USA first re-introduced the problems that motivated the initiation of the Cross-Cultural Cognition Multinational Project, with multinational cooperation in technology, commerce, and peacekeeping around the globe. Dr. Klein then went through the objectives, research methods, validity, and concerns before presenting details that need team members from all countries to discuss (see Appendix 3).

Discussion-Issues and Conclusions related to the predictor measures

1. An order effect was found in the framed line test (FLT) for the absolute judgment and also in the similarity task. The conclusion is that no further action is necessary unless there is an interaction with country.

2. The results from the US sample showed a bimodal distribution for the absolute judgment in the FLT. The conclusion is to re-analyze the data with the outliers (e.g., 2.5 SD) deleted in all four countries.
3. Categorization and similarity judgment: the results showed that the participants categorized based on the rule and used resemblance in similarity judgment across all four nations without a significant interaction with nation. This finding contradicts with the previous results (Norenzayan, et al, 2002). One suggestion from both the Japan and Taiwan teams is to exclude this test in future research, for the reasons that this test may be testing individual differences in thinking style or working memory capacity and is not sensitive to cultural differences. Another suggestion by Dr. Tae-Woo Park is to test it under time pressure. Given that the field battery must be tested in a short duration (estimated to be less than half an hour), the conclusion is to exclude this test unless new members are interested in using the test.
4. Facial expression test: the findings are somewhat different from the previous results with similar result patterns between Korea and USA and similar patterns between Japan and Taiwan. Three concerns were raised by Japan and Korea teams: (a) the use of half of the stimuli as used in the previous study, (b) the confound between social factors and emotion in rating, and (c) the method that may best capture the trend shown in the results among the four countries. Another concern was raised by the Taiwan team, as their analysis showed somewhat different result patterns. The use of Shakiness as suggested by the Japan team captures judgment variability which may be the first general indicator of the impact of background on judgment. Double checking of the results was recommended, given that the Taiwan team showed different results from their analysis. The conclusions are first to find a better

way to test the trend because the trends in different conditions showed different patterns in terms of the similarity in findings between countries, and second to run one experiment with only the neutral central face because this condition may be the most sensitive one to cultural differences.

Discussion-Issues and Conclusions related to the criterion measures

1. Attribution complexity scale: there was no significant interaction between the scale scores and country. Discussion focused on the construct validity and the underlying latent factor of the scale. Although the data obtained from Phase I were not satisfactory, Dr. Rue-Ling Chu from Taiwan showed some of her work with this scale which suggested that only one factor underlies the instrument across all the subscales and also that the scale is valid in testing individual differences in collectivesm. Dr. Incheol Choi from Korea discussed his research with the holism scale. No conclusion was made, but the inclination is to replace the complexity scale with the holism scale.
2. Exclusion test: the findings are consistent with previous results. Dr. Rick Warren from USA raised the issue on the cultural differences in the items that were excluded. Although previous studies did not show any differences in terms of excluding situational or disposition items, all team members agreed to send a photo copy of the raw data to Dr. Warren for his further analysis. Dr. Helen Klein also raised the concern on using such an extreme scenario. Dr. Yunn-Wen Lien from Taiwan described her work on attribution, which suggests that social norms play an important role in attribution. After a lengthy discussion, the conclusion is to test information exclusion with a Category x Consequence design. Category include

natural diasters (e.g., flooding, forrest fire), human caused accidents (e.g., train running into a department store, product failures), and unusal behaviors done by an animal or a person. The degree of severity would be used for the consequence factor.

3. Syllogism task: the results contradicts with previous findings. Dr. Lien from the Taiwan team first raised the issue over the results because their analysis showed somewhat different patterns. A lenghty discussion focused on the appropriate way to analyze the data by partialing out the individual differences in logic reasoning. The conclusion is to re-analyze the data by matching participants' logic ability in the abstract task condition across all four countries before conducting the statistical analysis on the data of the concret task condition. Also, the analysis can be conducted on a single index with correct rate from both the valid and invalid arguments. Finally, correlation in scroes between the abstract and concrete arguments should be computed.
4. The correlations among measures were rather weak, suggesting that more than one theoretical construct underlies the measures. The conclusion is to hold on this issue until further data analysis is completed.

Discussion- Issues related to the project

Many issues remain for the Rosetta project. Team members raised the following questions and issues.

1. The use of computer in future data collection.
2. How may phases will there be? How many countries will be included?
3. When will the project end? What is the final goal of the project?

4. Which among the old measures should be selected and what new measures should be added in the next phase?
5. The relations between all the measures should be hypothesized before rather than after the data collection. A lengthy discussion was held on the theoretical framework that encompasses all the measures and what "culture" truly means in social cognition.
6. Which component measures are most differentiated among different cultures?
7. How to improve communication? Telecommunication has been proved faulty, and workshop with experts in other fields related to cultural differences may be invited to provide feedback.
8. The time line for submiting the final report.
9. Issues to be discussed on Day 2 before the invited observers join the group.

There were two important conclusions. First, there will be only one final report for all teams and Dr. Helen Klein will be the person in charge. Second, a workshop will be held in Spring 2006. Yet, there is no conclusion in terms of the location of the workshop and the experts to be invited.

The meeting adjorned at 5 PM. Drs. Choi, Klein, Lien, and Radford returned to National Taiwan University to work on data analysis for the discussion on Day 2.

4 November, the 1st part (9 AM – 3:30 PM)

The workshop began with discussion on the new results from the further analysis. The basic findings on the syllogism test were similar even after the participants were

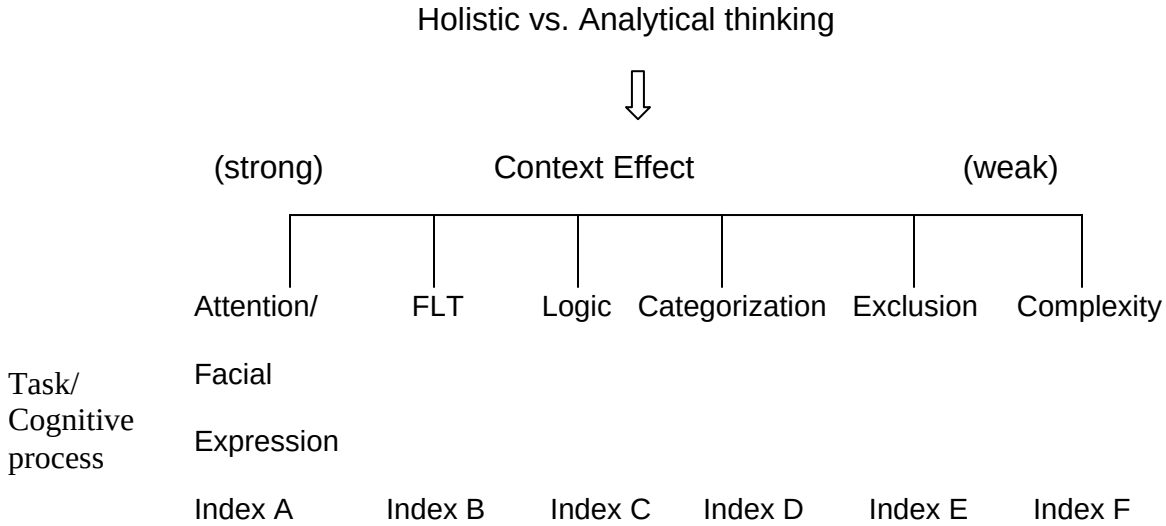
matched on the logic ability (see Appendix 4). The new analysis on the trends in the facial expression task was considered inappropriate because the background variable was not on an interval scale.

Discussion contined on the next workshop, what culture means, and the measures. A lenghty discussion focused on the definition of culture. Yet, this is a rather complex issue depending on the theoretical perspective a researcher holds.

Discussion related to the measures, the definition of culture, and Phase I project

1. Robustness of the measures. Attribution scale and the FLT test are the ones that showed reliable cultural differences.
2. The definition of culture. This is a rather complex issue. Who are the Westerners? Who is the Easterners? Where do culture differences come from? Are researchers testing “culture” or individual differences in thinking style?
3. What remains to be done for Phase I.
    A. The possibility of standardizing all the measures so that a single index to measure “holistic thinking” can be derived.
    B. The possibility of conducting a discriminant analysis to separate the Eastern and Western countries and also among the three Asian countries. Another variable should be considered is the education background. The participants with science and non-science background may differ significantly in thinking style.
    C. The effect size of each measure.

4. The expectation of new participants. Dr. Klein would like to get their feedback on the first phase, and other team members suggest that the new participants in Phase II can select the tasks from the battery and also include their choice of new tasks.
5. The access of data set beyond Rosetta. The suggestion is to design a mechanism with standard operational procedure for researchers to use the data set. This mechanism could begin with an e-mail alert to all members about the research issue to be addressed. Members who are interested in collaboration on the issue should respond within four weeks. A list of topics that have been addressed should be kept on the website so that redundancy can be avoided.
6. The theoretical framework for organizing the measures suggested by Dr. Lien.



*index: difference in the “context effect”

Dr. Lien also provided an example on the coding of the context effect.

4 November, the 2nd part (after 3:30 PM)

Two invited observers, Drs. Bhal and Khalid joined the workshop. Dr. Klein first introduced the project including the initial objectives, the outcomes from Phase I, and the purpose of extensions beyond the four countries. Each observer then gave a

presentation of their own research (see Appendix 4). The meeting adjourned with a discussion on the agenda for the last day of the workshop.

5 November

The workshop continued with Dr. Bhal's presentation and her questions on the project. The team members of Phase I (Dr. Masuda, Dr. Choi, Dr. Lien) presented each measure used in the battery and also the results from Phase I. Dr. Boff gave a presentation on the ultimate goals of the Rosetta project. The primary goal is to develop a robust tool that takes about 15 minutes to administrate for measuring cultural differences in cognition for culturally-sensitive human-system designs. Further discussion was on Phase II research.

Important issues related to the project conducted in Phase I

1. The final report is due on December 1. Dr. Klein will send a preliminary copy with new results and other team members should respond before the end of November.
2. Dr. Klein is in charge of the first publication based on the data set, and she may include graduate students who worked on the project. After the first paper, all team members are entitled to initiate an individual research paper based on the data set.
3. Any team member who wants to initiate an individual research paper should send an e-mail alert on the issue addressed. Other members have four weeks to respond for their willingness to participate in the collaboration. The initiator would be the first author and is responsible for the data analysis. The first author decides the order of authorship based on the contribution of the collaborators.

4. It is not necessary to acknowledge the support from AFOSR/AOARD in individual publication if conflicts arise. Yet, the acknowledgement is encouraged.
5. Teams that participate in Phase II project should submit the final research proposal no later than October, 2006.

Appendix 1

The Howard International House
Room 202
Taipei, Taiwan
November 3-5, 2005

This workshop is sponsored by
Air Force Office of Scientific Research,
Asian Office of Aerospace Research and Development

The Second Rosetta Workshop

Day 1, November 3 (Thursday)

◈ 9:00 – 9:10
Welcome & Introduction (Taiwan Team, Dr. Park, and Dr. Boff)

◈ 9:10 – 10:10
Discussion of Original Objectives, Goals, & Expectation (Prof. Klein)

✦ 10:10–10:30 (Coffee break)

◈ 10:30 -12:00
- **Discussion of results: National differences, interrelationships of perception and cognition**
  - What have we learned?
  - Further analysis and modifications

✦ 12:00 – 13:30 (Lunch) *LA MODE CAFÉ (B1)*

◈ 13:30 – 15:00
- **Discussion of methods: Procedures, participants (including demographics), and materials**
  - What have we learned?
  - Equivalent vs. identical procedures?
  - Needed changes: omissions, modifications & additions

✦ 15:00 – 15:30 (Coffee break)

◈15:30 – 17:00
- Phase I research process: Coordination, data sharing, & communication.
- Lessons learned
- How could we have improved these?

◈ 18:00 (Meet at front door)

✦ **19:00 (Working dinner: Continue discussions)**
The Landies-Tien Hsiang Lo

The Second Rosetta Workshop

Day 2, November 4 (Friday)

◈ 9:00 – 10:10

- **Current & future publications presentations (Dr. Warren)**
  - Authorship guidelines
  - Final report
  - Consolidated research paper
  - Individual research paper

✦ 10:10 – 10:30 (Coffee break)

◈ 10:30 – 12:00

Time line for the final report (Dr. Park)

✦ 12:00 – 13:30 (Lunch) ***GARDEN CAFETERIA (F1)***

◈ 13:30 – 15:00

- **Discussion of the Phase II research process: Coordination, data sharing, & communication (Prof. Klein)**
  - Perspectives & concerns of Phase I Teams
  - Review of original goals & objectives: How can we improve our work?

✦ 15:00 – 15:30 (Coffee break)

◈ 15:30 – 17:00

Welcome Phase II team - Introductions by entire team (Dr. Park)

◈ 15:45 – 17:00

- **Summary Overviews and Perspectives & Issues (Prof. Klein)**
  - Summary Overview or Current Rosetta Project
  - Malaysia Perspective & Issues of Rosetta Project (Prof. Halimahtun)
  - India Perspective & Issues of Rosetta Project (Prof.Bhal)

✦ **17:30 - 18:30 (NTU tour)… optional**

✦ **19:00 (Dinner)** ***The Howard Plaza Hotel -Formosa***
The Second Rosetta Workshop
Day 3, November 5 (Saturday)

◈ **9:00 – 10:10**

- **Cultural Research in Applied Contexts (Prof. Klein)**
- **Cultural Research in Laboratory Contexts (Dr. Choi)**
- **The Rosetta project (Prof. Klein)**

✦ **10:10 – 10:30 (Coffee break)**

◈ **10:30 -12:00**

- **The research questions: Perception and Cognition (Prof. Klein)**
- **Research Methods: The test battery, participants, and procedures**
- **Outcomes and remaining questions.**

✦ **12:00 – 13:30 (Lunch)** ***LA MODE CAFÉ (B1)***

◈ **13:30 – 15:00**

- **The second phase of the project (Prof. Klein)**
- **Extending the research tools: Better answers to the research questions.**

— How can we accommodate national differences without compromising outcomes?
— Can we introduce more naturalistic decision making scenarios?
— Can we include an assessment on attention mechanisms?
— Inclusion of short personality and cultural assessment scales?
— Extending the samples: Broader understanding of cultural characteristics.

- **Theoretical and practical advantages**
- **How can we make this work well for all partners?**

✦ **15:00 – 15:30 (Coffee break)**

◈ **15:30 – 17:00**
Future research plans & Closing (Dr. Boff)

The Second Rosetta Workshop

Participants

Dr. Halimahtun M. Khalid
Dr. Helen Klein
Dr. Ken Boff
Dr. Rik Warren
Dr. Mark Radford
Dr. Incheol Choi
Dr. Tae-Woo Park
Dr. Terry Lyons
Dr. Takahiko Masuda
Dr. Kanika Bhal
Dr. Yunn-Wen Lien
Dr. Cathy Chu
Dr. Yei-Yu Yeh

Staff
Joseph Wen
Ann Yang
Judy Weng
Wei-Chien Wang

Appendix 2

Preliminary results

Rosetta Results: Participants
September 2005

I. PARTICIPANTS.

Four samples of undergraduate students served as participants in this study, See Table I for the demographic characteristics of the groups. The participants for the samples were selected from Hokkaido University in Japan (N = 94), Seoul National University in Korea (N = 92), National Taiwan University in Taiwan (N = 99), and Wright State University in the United States (N = 94). Japan, Korea, and Taiwan are East Asian groups and likely to include a preponderance of holistic thinkers. The U.S. is Western and likely to include more analytic thinkers. There is strong research establishing the holistic thinking patterns of Japan, Korea, and Taiwan and the analytic thinking patterns of the U.S.

All participants were undergraduates enrolled in a course in Introductory Psychology at their university. They completed the study as part of a course requirement. Potential participants were included only if they report that their parents were native to Japan, Korea, Taiwan, or the United States, respectively, and that they had not lived away for more than one (1) year.

Age. Participants were between 17 to 24 years of age. The mean ages of the samples from Japan, Korea, Taiwan, and the U.S. were 19.54, 20.99, 19.75, and 18.89, respectively. These ages differed significantly ($F = 48.94$, $p < 0.001$). The Korean sample had the highest mean age and the U.S. the youngest. See Table I for the standard deviations of ages.

Variable	Japan N = 94 %		Korea N = 92 %		Taiwan N = 99 %		United States N = 94 %	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age	19.54	1.03	20.99	1.35	19.75	1.32	18.89	1.08

Gender. No attempt was made to equate the numbers of males and females in each sample. The percent of males were 60.6%, 42.4%, 22.2%, and 21.3% respectively. Gender differences were significant over the four groups ($\chi^2 = 43.16$, $df = 3$, $p < 0.001$) with the Japanese sample having the most males and the U.S. sample the least.

Family Background. The demographic data supported the placement of participants to the groups. All had parents who were native to the group. All participants from Japan, Korea, and the U.S. spoke Japanese, Korean, and English, respectively, as their first language. All participants from the Taiwan group spoke Chinese or Taiwanese as their first language. There were some differences in the educational levels of parents.

Academic major. Because of differences in degree requirements and student choices, the majors of students enrolled in Introductory Psychology varied significantly over samples. (χ^2 = 245.54, df = 24, $p < 0.001$) Among the Japanese participants, the most frequent majors were Social/Behavioral Sciences (36.2%), Humanities/Fine Arts (39.4%), and Law (12.8%). For the Korean group, the most common majors were Engineering (24.4%), Health Sciences (17.8%), Humanities (11.1%), and Natural Sciences (11.1%). Among the Taiwan participants, 41.4% majored in Social /Behavioral Science while 24.2% majored in business. There were 12.1% in both Natural Science and Humanities/Fine Arts. For the U. S. participants 26.6% were majoring in Health Sciences, and 19.1% in both Social/Behavioral Sciences and Education.

	Japan	Korea	Taiwan	United States
Variables	N = 94	N = 92	N = 99	N = 94
	%	%	%	%
Engineering	3.2	24.4	2.0	5.3
Social/Beh Science	36.2	4.4	41.4	19.1
Natural Science	1.1	11.1	12.1	3.2
Business	0.0	10.0	24.2	7.4
Humanities/Fine Arts	39.4	11.1	12.1	9.6
Education	7.4	3.3	0.0	19.1
Health Sciences	0.0	17.8	7.1	26.6
Law	12.8	2.2	1.0	1.1
Other	0.0	15.6	0.0	8.5

Year in School. Degree requirements and student choices also effected when students enrolled in Introductory Psychology courses. The samples differed significantly in the distribution of students over year in school. (χ^2 = 76.59, df = 9, $p < 0.001$) In the Japanese group, 66% were in their first year of study, 25.5 % in their second year, and 8.5% in their third year. In the Korean group, 58.7% were in their first year of study, 25% in their second year, 12% in their third, and the remaining 4.3% in their final year. In this sample from Taiwan, 26.3% were in their 1st year of study, 34.3% were in their 2nd year, 18.2% in their third year and the remaining 21.2% in their final year. Finally, in the U. S. sample, there were 74.5% were 1st year students, 18.1% 2nd year, 6.4% 3rd year and 1.1% 4th year.

II. THE MEASURES AND THEIR DIFFERENCES BY NATIONAL SAMPLE

Order Effects.

This study was designed to run in blocks counterbalanced for the orders of the measures within the first and the second days of assessment. Because of incomplete

data, the final samples included unequal numbers of participants in the blocks. An Analysis of Variance queried the significance of the differences over samples introduced by presentation order.

First Day.

The first day of testing included the three criterion tasks administered in a group session. The tasks were presented in six counterbalanced orders. Performance for the first day tasks showed no significant order effects for measures.

Second Day.

The second day of testing included the three predictor tasks administered in an individual session. The Framed Line Task had two orders with the relative judgment first in one and absolute in the other. For Similarity –Belonging, half of the participants were given the Similarity Task and the other half the Belonging condition. This made up twelve different orders. Performance for the second day showed significant order effects for the absolute judgment of the Framed Line Test (F=1.89, p=. 038) and Similarity Task (F= 2.42, p = .01).

An analysis of all participants found significant FLT order effects for FLT for performance on the absolute condition. Those participants who received the relative task first did better than those who received the absolute task first (F = 8.83, p = 0.003).

Framed Line Test: Presentation Order					
	Absolute Task First		Relative Task First		
	Mean	SD	Mean	SD	F
Absolute judgment Error scores	48.08	28.04	40.15	23.86	8.83**
Relative judgment Error scores	29.22	21.54	26.36	16.23	2.15

** p < .01

The FLT order effect was then examined in each sample. The order effect was not significant for the Japanese sample (F = .28, p = .60) and Taiwan sample (F = 3.34, p = .07) but was significant for the Korea sample (F = 8.09, p = .006) and the U.S. sample (F = 6.17, p = .015).

Criterion Measures: Cognitive

Exclusion Attribution Scale: "The Murder Mystery"

The Exclusion Attribution assessment taps Analytic vs. Holistic reasoning. Analytic reasoning was expected to lead to the exclusion of more items that would holistic reasoning. This is because attribution is focused on the dispositional rather than being inclusive of the situational. Participants were first presented with a scenario. They were then asked which of a list of 97 information items they would exclude as irrelevant for making a decision about the scenario. It was predicted that the U.S. sample, as a hypothesized analytic sample, would exclude more items than would each of the three, presumably holistic samples.

Mean exclusion rates (number of items) were 42.29, 39.04, 40.18, and 51.06 for the Japanese, Korean, Taiwanese, and U.S. samples respectively. The four samples differed overall, ($F=11.56$, $p < 0.001$). The U.S. sample excluded more items than Japan ($t =3.74$, $p < 0.001$), Korea ($t = 5.44$, $p < 0.001$), and Taiwan ($t = 4.89$, $p < 0.001$). The lower exclusion rates for the three Holistic samples, Japan, Korea, and Taiwan samples, supports the research hypotheses that holistic reasoning is associated with the incorporation of a wider range of information. The three East Asian samples did not differ from each other.

Conclusions and Concerns

The results were consistent with the earlier work of Choi, Dalal, Kim-Prieto, & Park, (2003). They also found marked differences between the exclusion rates of Far Eastern and Western participants consistent with differences in holistic and analytic thinking.

This finding is particularly impressive because the scenario is so specific. Do we think that it would be effective with non-student samples –business people, airline pilots, etc?

Attribution Complexity Scale: "Agree or Disagree?"

The Attribution Complexity Scale assesses the complexity of attributions using a self-report scale (Fletcher, Danilovics, Fernandez, Peterson, & Reeder, 1986). Based on this earlier research, analytic thinkers were expected to show lower attribution complexity than were holistic thinkers. The Scale includes seven subscales focused on specific components. These were: Level of interest or motivation (MOT), Preference for complex explanation (PCE), Presence of metacognition concerning explanation (MET), Behavior as a function of interaction (BFI), Complex internal explanation (CIE), Complex contemporary external explanation (CCE), and Tendency to infer external causes operating from the past (TEM). Each of the seven was designed to reflect an aspect of attribution complexity.

Combined Scores.

We first looked at the combined score using all seven (7) scales. The mean scores for Japan, Korea, Taiwan, and the U.S. were 141.05, 144.52, 143.96, and 144.46 respectively. These are not significantly different and therefore provide no support for the hypotheses of sample differences on combined scores as a measure of attribution complexity.

The combined score had a high reliability coefficient overall (α = .86). In addition, all four samples had high reliability coefficients, Japan (α = .88), Korea (α = .84), Taiwan (α = .86), and the U.S. (α = .89), indicating reliability of the Attribution Complexity Scale.

Subscale Scores

A *post hoc* analysis then looked for national differences for the seven subscales that make up the Attribution Complexity Scale. See Table below for mean scores and standard deviations by samples as well as F values and significance levels for sample differences.

Attribution Complexity Scale Performance

Variables	Japan N = 94		Korea N = 92		Taiwan N = 99		United States N = 94		
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	F
ACS Scores									
	141.05	19.57	144.52	15.30	143.96	17.06	144.46	19.96	
AC Subscales									
Motivation (MOT)	20.27	3.99	18.76	4.36	21.01	3.75		4.42	5.23*
Preference (PCE)	18.39	3.73	19.80	3.58	17.20	4.26	20.56	3.79	8.03**
Metacognition (MTC)	19.61	4.44	21.08	3.40	20.95	3.36	19.14	3.38	4.85*
Behavior (BFI)	21.32	3.50	20.85	2.26	22.63	2.80	21.55	4.15	5.16*
Internal (CIE)	22.36	2.87	22.40	2.84	20.94	3.11	21.52	3.25	10.17**

Contemporary CCE)	19.66	3.92	20.36	3.43	20.53	3.42	20.45	3.62	2.48
Past Orientation (TEM)	19.45	4.01	21.33	3.25	20.71	3.64	21.11 20.13	3.79	4.44*

* p < .05
** p < .01

The original research with U.S. participants found the seven scales to be moderately and positively correlated. Correlation matrices for each of the four samples showed that most inter-item correlations were highly significant. These tables are available as an htm file “ACSubscales.htm” in the Sept05 Web site folder.

Conclusions and Concerns.

The combined score of the Attribution Complexity Scale did not show expected differences for the groups. An analysis of subscales showed differences between the samples for six of the seven measures. These do not follow a predicted pattern or a discernable one. Additional research would be needed to identify patterns in these relationships.

How can we interpret this pattern of responses? It would be most interesting to identify group differences in the nature of complex attribution.

Syllogisms Task: Is it Logical?

The 'Is it Logical' test asked participants to judge the argument conclusions for a set of syllogisms. The syllogisms varied by their logic (valid and invalid) and by the believability (believable vs. non-believable) of their conclusions. Consistent with earlier work (Norenzayan et al., 2002) it was expected that analytic thinkers would favor the formal rules of logic over the intuitive appeal of believability in responding to argument conclusions. In contrast, the holistic thinkers, relying more on intuition, were expected to respond more to the believability of argument conclusions.

In order to interpret the outcomes, two checks were included and are reported first. To confirm the assumed believability of the conclusion statements, participants rated the believability of the argument conclusions alone. To assess the ability of the participants to attend to logic in the absence of competing factual information, abstract syllogisms were included. These abstract syllogisms used letters and nonsense words to present valid and invalid argument conclusions. Performance on these abstract syllogisms assesses logic independent of believability.

Believability of conclusions.

To confirm the assumptions about believability, participants rated the believable of each of the argument conclusion on a scale from -3 (Definitely False) to +3 (Definitely True). For each sample, we averaged the responses across all eight (8) believable and across all eight (8) non-non-believable statements. A mean value greater than zero for the 'believable' statements indicates that participants believed the conclusions to be true. A mean value less than zero indicated that participants believed the conclusions to be untrue. The means of ratings by believable and sample are provided in the table below.

Believability Check

Variables	Japan N = 94		Korea N = 92		Taiwan N = 99		United States N = 94	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Believable Conclusions	.92	.68	.96	.64	1.59	.57	1.93	.63
NonBelievable Conclusions	-2.09	.55	-2.05	.60	-2.16	.69	-1.87	.64

We evaluated the ratings for both believable and non-believable items for each of the samples. All comparisons were significant at the $p < .001$ level. Sample values appear below:

- The Japanese sample gave the believable conclusions a mean rating of M = .92. This is significant ($t(93) = 13.20$, $p < .001$). Their mean rating of the non-believable conclusions was also significant M = -2.09; $t(92) = -36.85$, $p < .001$).

- For the Korean sample, believable conclusions given a mean rating of M = 0.96 (t (91) =14.41, p < .001). The non-believable conclusions had a mean rating of -2.05, (t (91)= -32.87, p < .001).

- For the Taiwan sample, the believable conclusions had a mean rating of M = 1.59 (t (98) =27.78, p < .001). The non-believable conclusions received mean ratings of -2.16, (t (98)= -31.09, p < .001).

- The U.S. sample gave the believable conclusions a mean rating of M = 1.93, (t (93) =29.91, p < .001). Non-believable conclusions were given a mean rating of M = -1.87 (t (93)= -28.29, p < .001).

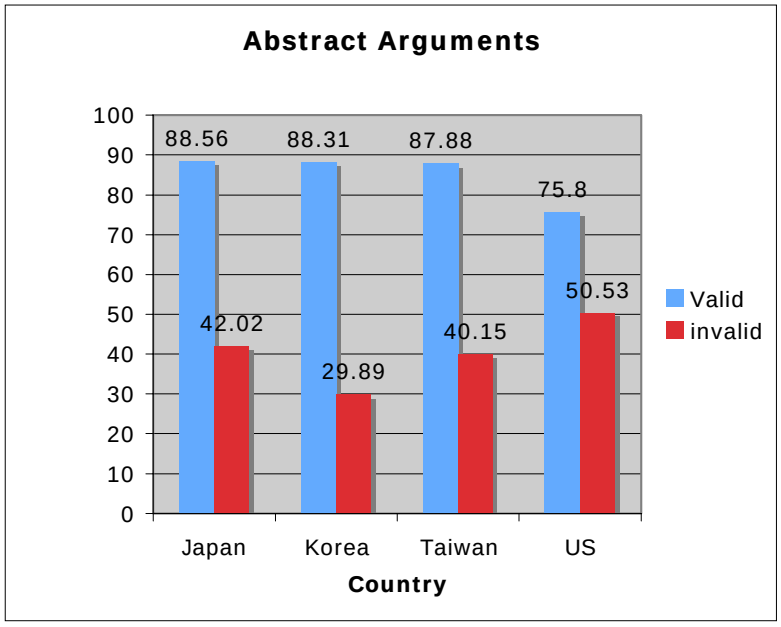
The believable check confirms the classification of items for these samples. All four samples, rated the believable conclusions to be significantly higher than zero and the non-believable one as significantly less than zero. We then looked to sample differences in judgments of the believable and non-believable statements. We found overall differences among the samples in judgments of believable conclusions (F = 57.86, p < .001). *Post hoc* analysis showed that, the U.S. sample found believable conclusions to be significantly more believable than did the Japan, Korea and Taiwan samples, (t = -10.64, p < .001, t = -10.42, p < .001, and t = -3.95, p < .001, respectively). The Taiwan sample found believable conclusions to be significantly more believable than Japanese and Korean samples, (t = -7.47, p < .001, and t = -7.16, p < .001, respectively), and less believable than the U.S. sample.

Analyses showed that there were differences between the samples in the non-believable conclusions as well, (F = 3.76, p = .011). In a *Post hoc* analysis, the Taiwan and U.S samples differed in their rating of the non-believable conclusions (t = -3.01, p = .003) with the Taiwan participants rating them as less believable Although the participants believed the 'believable' conclusions, the samples differed in the magnitude of their judgments. A covariate analysis addressing this potential problem is described the Confounding Effects section concern below.

Logical Ability: Abstract Arguments

The abstract syllogisms assessed logical processes independent of differences in believability. We looked at performance differences in logical ability in three ways:

Pattern of Response. We first looked at the general pattern of response. Participants indicated that they thought that the believable conclusions to be valid, indicating 'yes', for 88.56%, 88.31%, 87.88%, and 75.80% of items for Japan, Korea, Taiwan, and the U.S samples, respectively. The percentages of 'yes' responses indicating an non-believable statement as valid were 42.02%, 29.89%, 40.15%, and 50.53% for Japan, Korea, Taiwan, and the U.S sample, respectively. All samples correctly rated the believable arguments as valid greater than chance (i.e. 50%). All samples, except the U.S sample, rated invalid arguments correctly and better than chance. Values were Japan (57.98%), Korea (70.11%), Taiwan (59.85%), and the U.S (49.47%). The figure below, labeled Abstract Arguments, shows the pattern of responses on valid arguments and invalid arguments. Quantitative analyses of the differences follow.



Response Bias. We then evaluated response bias, the tendency to respond 'yes', for the abstract items. See table below. Because four items were actually valid and four were not, an accurate participant would respond with four (4) 'Yes's and four (4) 'No's'. A deviation from 50% would indicate response bias. The mean rates of responding 'Yes' for Japan, Korea, Taiwan, and the U.S samples were 65.29%, 59.10%, 64.02%, and 63.16%, respectively. Overall, the participants showed a significant response bias ($F(3, 375) = 3.73$, $p = 0.012$) with all samples responding 'Yes' above 50%. A post hoc analysis showed that the Japanese were more likely to respond 'Yes' than the Koreans ($t = 3.84$, $p < .001$). Other sample differences were not significant. These differences in response biases need to be considered in interpreting the outcomes from the concrete arguments.

Abstract Arguments

Variables	Japan N = 94		Korea N = 92		Taiwan N = 99		United States N = 94	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Mean % saying 'yes'	65.29	12.32	59.10	9.47	64.02	12.15	63.16	17.99
% Saying 'yes' to valid	88.56	14.97	88.32	16.76	87.88	16.12	75.80	23.60
% Saying 'yes' to invalid	42.02	17.28	29.89	11.27	40.15	18.15	50.53	20.73
Accuracy	46.54	20.93	58.42	21.38	47.73	24.25	25.27	26.05

Accuracy. To evaluate performance in discriminating valid from invalid arguments while controlling for response bias, a single measure of accuracy was computed: hits (% of 'Yes' responses for valid arguments) minus false alarms (% of 'Yes' responses for invalid arguments). The mean accuracy scores for Japan, Korea, Taiwan, and the U.S samples were 46.54%, 58.42%, 47.73%, and 25.27%, respectively. These accuracy differences are significant ($F(3, 375) = 33.26$, $p < .001$).

Post hoc analysis showed that Japanese, Korean, and Taiwanese samples were more accurate than the U.S. sample ($t = 6.17$, $p < .001$, $t = 9.48$, $p < .001$, and $t = 6.20$, $p < .001$, respectively). *Post hoc* analysis also showed that the Korean sample was more accurate than the Japan and Taiwan samples, ($t = 3.83$, $p <. 001$ and $t = 3.22$, $p = .001$, respectively). The samples differ in accuracy complicating interpretations of the concrete arguments. The covariate analysis below addresses this concern.

Concrete Arguments

Each participant judged syllogisms that varied in logic and believability. Sample differences are given in the table below. Based on differences in Analytic – Holistic reasoning, we anticipated that the East Asians, relative to the U.S sample, would be more likely to evaluate argument as valid when the conclusion is believable, and less likely to do so when the conclusion is non believable.

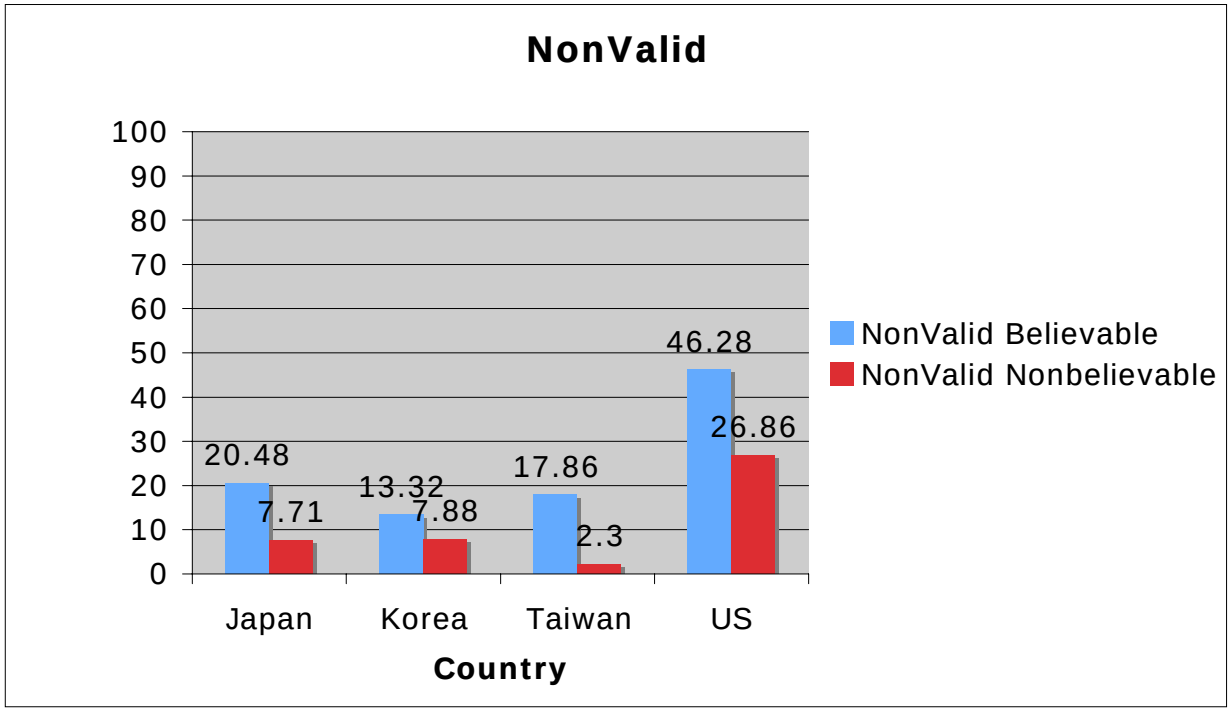
Concrete Arguments

Variables	Japan N = 94		Korea N = 92		Taiwan N = 99		United States N = 94	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Valid Believable	91.76	12.90	83.97	18.38	94.13	12.84	93.09	14.39
Valid Non-Believable	20.48	19.04	13.32	15.48	17.86	18.30	46.28	22.88
Invalid Believable	88.56	23.66	82.07	19.73	77.55	26.50	66.49	35.08
Invalid Non-Believable	7.71	12.71	7.88	15.25	2.30	8.10	26.86	19.82

A Sample (Japan, Korea, Taiwan, U.S.) by Argument Validity (valid vs. invalid) by Conclusion Believability (believable vs. non-believable) ANOVA tested this hypothesis.

There was a main effect of Sample, $F (3, 374) = 22.77$, $p < .001$, indicating differences among national samples. The participants in the East Asian samples out-performed those in the U.S. sample. There was a main effect of Argument Validity, $F (1, 374) = 3463.60$, $p < .001$, indicating sensitivity to logical structure. Participants rated valid arguments as more valid than invalid argument. There was also a main effect of Conclusion Believability, $F (1, 374) = 163.51$, $p < .001$, indicating that belief bias influenced judgment.

Finally, the analysis showed a significant three-way interaction between sample, conclusion argument validity, and believability, $F (3, 374) = 5.35$, $p <. 001$**.** The Figure below describes this interaction. The U.S. sample showed larger difference between believable and non-believable, valid and invalid syllogisms than did the other three samples. The direction of this effect, however, refutes rather than supports the expectation.



Confounding effects.

We were concerned with the potentially confounding effects of several variables. First, because the samples differed in their assessment of the 'believability' of the argument conclusions, it might be possible that believability would confound performance. Second, because we found differences in logical ability as measures with the abstract logic item analysis we were concerned that this would also differentially influence performance. Next, we were concerned with the differences noted in academic majors among the samples. Perhaps, for example, students majoring in the physical sciences and engineering received more training in logic and also varied in distribution over the four samples. The majors varied over the samples. Two additional demographic variables, age, and gender might also confound performance.

A covariate analyses assessed the potentially confounding effects of differences in judged believability, logical ability, academic major, age, and gender. Major was dichotomized into science vs. non-science (physical sciences, engineering, vs. social sciences and the humanities). When each variable was included independently in analysis, the main effects of sample, logic (logical vs. non-logical), and believability (believable vs. non-believable) remained significant and the interactions remained significant. There was no evidence that any of the five variables altered the results. When the five possible covariates were entered together, the 3-way interaction was no longer significant but all others remained significant.

Conclusions and concerns.

The outcomes from the Logic Task did not support the predicted relationships between analytic vs. holistic reasoning and performance on the syllogisms. Two explanations might be suggested for this.

Sample Differences. The U.S. university is less selective than the other three and the level of the students may have contributed more variance than did group differences in cognition.

Procedural Bias. The second explanation rests with the format used to present the syllogisms during testing. The format used in presenting the syllogisms may have been one less common to U.S. students. In this study, participants were asked to indicate if a syllogism in the following format was logical:

> Premise: [Statement]
> Premise: [Statement]
> Conclusion: [Statement]

As was discussed at during the planning session, U.S. students may have been more likely to encounter the format:

> If: [Statement]
> And If: [Statement]
> Then: [Statement]

This was the format used during the U.S. pilot testing. The team made the decision to use a single format. The unexpected outcomes might be attributed to the unfamiliar format. To achieve a more context sensitive measurement of logical processes, we may need to provide participants in each group with the format that can optimize the use of logic. Because our goal is to measuring underlying cognitive processes rather than task familiarity, this means using the format familiar to the group to be assessed. This would be particularly important as the protocol is used with a broader range groups. We may want to review this distinction during our meetings

Are there any other potential explanations?

Predictor Measures: Perceptual

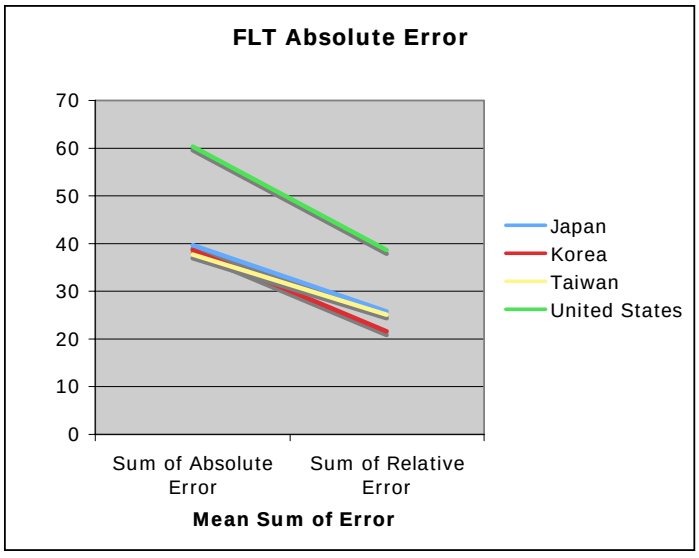
Framed Line Test

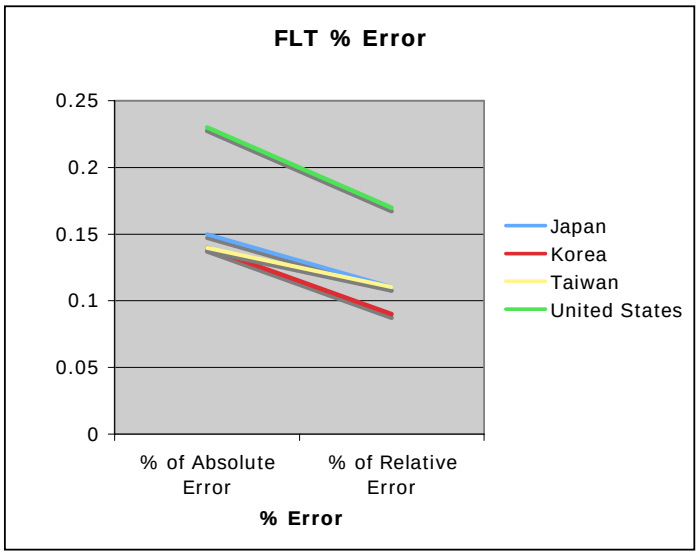
The Framed Line Test measures accuracy in judging an absolute or relative length presented in one framing square and reproducing it in a second framing square. We expected the more holistic Japanese, Korean, and Taiwan samples, to do better than the more analytic U.S. sample in the relative condition because they would be attuned to the broader context of the judgment. We expected the more analytic U. S. sample to do better than the East Asian samples in the absolute condition because they focus on the line alone. We hypothesized an interaction between the samples and the condition, relative vs. absolute judgment.

Absolute vs. Relative Performance.

We first looked at performance on the two conditions. Performance using an absolute error score as well as was a percent error score. See figures below.

For the measure absolute error score, the performance was better on the Relative condition than the Absolute condition for all samples, Japan ($t = -6.69$, $p < 0.001$), Korea ($t = -9.29$, $p < 0.001$), Taiwan ($t = -7.35$, $p < 0.001$), and the U.S ($t = -5.13$, $p < 0.001$). For the percentage error measure, the Relative condition performance was better than that on the Absolute condition for all samples, Japan ($t = -5.29$, $p < 0.001$), Korea ($t = -7.48$, $p < 0.001$), Taiwan ($t = -4.48$, $p < 0.001$), and the U.S ($t = -3.26$, $p = 0.002$). The interaction between the sample and the condition was not significant.





The Relative Condition.

The table below details sample differences in performance. In the relative condition, the mean of Σerrors over the five trials for Japan (25.74mm), Korea (21.58 mm), Taiwan (25.08 mm), and the U.S. (38.66mm) samples were significantly different overall (F= 16.24, $p < 0.001$). The U.S. was significantly less accurate than the Japanese ($t = -3.80$, $p < 0.001$), Korean ($t = -5.65$, $p < 0.001$), and Taiwanese $t = -4.52$, $p < 0.001$) samples with equal variance not assumed. The three East Asian samples were not significantly different from each other. This outcome suggests the superiority of holistic groups for the relative task.

Framed Line Test Performance

Variables	Japan		Korea		Taiwan		United States	
	N = 94		N = 92		N = 99		N = 94	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Sum of Absolute Error	39.71	21.42	38.65	16.38	37.72	17.91	60.30	37.24
Sum of Relative Error	25.73	18.58	21.58	10.09	25.08	10.65	38.66	27.24
% Absolute Error	.15	.08	.14	.06	.14	.07	.23	.14
% Relative Error	.11	.08	.09	.04	.11	.05	.17	.12

The Absolute Condition

The analysis of the absolute data, however, complicates the interpretation of Framed Line Test outcomes. The variable of sample was significant overall (F= 18.26, $p < 0.001$). In the absolute condition, the mean of Σ errors over the five trials for Japan (39.71 mm), Korea (38.65 mm), and Taiwan (37.72 mm) was each less than for the U.S. (60.30 mm) sample. The U.S. mean was significantly less accurate than the Japanese (t

= -4.65, $p < 0.001$), Korean ($t = -5.11$, $p < 0.001$), and Taiwanese ($t = -5.41$, $p < 0.001$) samples. The three holistic samples were not significantly different from each other.

Conclusions and Concerns.

For both Relative and Absolute judgment conditions, the three East Asian groups were superior in performance to the U. S. sample. While his was expected in the Relative judgment condition, it is contrary to past findings for the Absolute judgment condition.

I (HAK) am uneasy with the high errors of the U.S. sample for the relative condition and more so for the absolute condition. The high error rates are accompanied by high standard deviations. A review of scatter plots suggests that a subset of participants contributed to the high error rates. This may be because they were unable to make absolute judgment or that they were unable to understand the directions provided.

During pilot tests in the U.S., prior to finalizing procedures, the instruction included a sample judgment. After completed a sample judgment, participant was guided as they measured their response and compared it to the correct response. While most of the participants provided fairly accurate judgments, the procedure insured that all participants understood the task. It may be possible that the training provided during the pilot testing was needed with this sample. If we want to insure that we are measuring underlying perceptual processes rather than task familiarity, perhaps a more robust instruction procedure would be appropriate. How might this be managed in future work?

Belonging and Similar

Analytic vs. holistic thinking has been associated with differences in categorization. The Similarity – Belonging Task suggested by Norenzayan, et al (2002), assesses differences in categorization. A request to classify by 'belonging' was expected to tap a rule-based cognitive strategy characteristic of analytic thinkers. While this should favor analytic thinking, past research has not found this difference with educated samples. The request for classification by 'similarity' was expected to tap holistic thinking: the more intuitive, exemplar-, or family-based strategy. This more holistic demand characteristic was expected to favor the East Asian samples. We measured both correct responses and time to completion for both categorization tasks. Each participant was assigned to only one of the conditions so that the sample sizes are smaller than for other measures.

Belonging.

For the Belonging condition, there were 49, 44, 49, and 48 participants in the samples from Japan, Korea, Taiwan, and the U.S. respectively. In the 'belonging' or rule–based condition, the average number of correct responses over all samples was 13.01 (SD = 4.91). The mean time to completion was 184.49 sec, SD = 64.47 sec. The table below shows means and times by samples. Neither score nor time to completion reached significance across the samples or in pair-wise comparisons.

Categorization Performance

Variables	Japan		Korea		Taiwan		United States	
	N = 94		N = 92		N = 99		N = 94	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Rule Condition	12.92	5.56	13.52	5.45	13.37	4.39	12.27	4.21
Similar Condition	10.65	4.34	11.44	3.63	11.98	2.79	11.54	3.53
Time: Rule	174.04	57.63	184.48	57.78	190.78	81.92	188.75	56.93
Time: Similar	172.00	51.68	179.34	67.01	184.16	80.27	229.09	64.46

Similar.

For the Similar condition, there were 43, 44, 50, and 46 participants in the samples from Japan, Korea, Taiwan, and the U.S. respectively. In the "similar' condition, the average number of correct responses over all samples was 11.43 (SD = 3.58). The mean time to completion was 191.64 sec. SD = 70.30 sec. Table below shows means and times by samples. While accuracy was not significantly different among the samples, ($F = 1.09$, $p = 0.36$), times to completion was differed ($F = 6.65$, $p < 0.001$). Time to completion was significantly longer for the U.S. participants compared to the Japanese, the Korean, and the Taiwanese participants ($t= 4.59$, $p < 0.001$, $t=3.53$, $p = 0.001$, and $t= 3.00$, $p= .003$, respectively). The U.S. participants were able to make the intuitive judgments but they took significantly longer to do so.

Conclusions and Concerns.

The outcomes from both conditions tend to support past research. No differences were found in the belonging condition with educated samples consistent with past work. Differences were found in time to completion for the Similar condition.

Similarity did show an order effect. Because of unequal numbers in cells, care should be taken in interpreting this.

Facial Expression

The Facial Expression task (Masuda, et al. 2005) was used to assess the participant's judgment of a face's emotional expression in the context of peripheral faces with the same or different emotions. We were interested in the relationship of sample, central figure expression, and background figure expression for the 'Happiness' rating and the 'Sadness' rating. East Asians participants, with more holistic, field dependent perception, might be more influenced by peripheral faces than would the more analytic, field independent U.S. participants. This was examined in two ways:

First, we expected sample differences in the saliency of the background figures: the East Asian samples were expected to report the background figures to be more salient and would also report that the background would be more likely to affect their judgment. Four questions queried the relationships of background salience and perceived impact.

Second, we expected that 'happiness' and 'sadness' judgments of central figures would be more affected by the background figures in the Japanese, Korean, and Taiwan samples than it would be in the U.S. sample. We also wanted to know if characteristics of the central figures – happy, neutral, or sad – would influence responses. There were three ways to look at facial expression performance. We looked at the facial expression judgments using three-way ANOVAs of Sample X Central Figure Emotion X Background Figure Emotion for 'happiness' and for 'sadness' judgments. We also looked at six two-way comparisons for central faces: the emotion of the central face (happy, neutral, or sad) and by the judgment (happiness or sadness). In each comparison, we looked at the main effect of sample, of background, and of sample X background interaction. See Table below for means for samples by background. Finally, we looked at the 'Shakiness' vs. Stability of the judgments.

Background salience and perceived impact.

To tap the salience of the background figures, participants indicated if they noticed the background changes. The table below shows percentage indicating noticeable change in background for each sample. There were significant differences ($\chi 2 = 24.66$, $p < .001$) in the reports for the four groups. The results of the Taiwan sample did not meet expectations. There was also a significant sample difference ($\chi 2 = 32.43$, $p < .001$) when participants were asked if the background affect their judgment of the central figure. See table. These are consistent with expectations based on Holistic versus Analytic differences.

Responses to Change in Background

Variables	Japan N = 94		Korea N = 92		Taiwan N = 99		United States N = 94	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Notice Change (Yes)	98.9%		90.2%		75.8%		86.2%	
Affected by Change (Yes)	74.5%		64.8%		57.6%		35.1%	
Same: Ease Judgment	3.68	2.54	5.17	2.18	4.21	2.30	4.30	1.91

Different: Judgment	Ease	6.79	2.34	5.74	2.31	6.58	2.15	7.06	2.32

Two questions queried the perceived impact of the background on judgments. When asked how easy/difficult it was to judge central figure when the background differed, there was significant different among the samples ($F = 4.52$, $p = 0.004$). The table provides mean values. The U.S sample did not differ from the three samples. Korea judgments of easiness were higher than those of the Japan ($t = 3.53$, $p < 0.001$). When asked how easy/difficult it was to judge the central figure when the background was the same, there was significant different ($F = 3.19$, $p = 0.024$). See table above. The U.S sample's judgments of easiness were higher than those of the Korean's ($t = -2.62$, $p = .01$). There were no differences between the other three East Asian samples.

Relationships among variables: Sample X Central Figure X Background Figures

The table below shows mean judgments for Happiness and Sadness judgments.

Facial Expression Judgments

Variables	Japan N = 94		Korea N = 92		Taiwan N = 99		United States N = 94	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Happiness Judgment								
CF: Happy BG: Happy	6.96	1.17	6.42	1.18	6.02	1.21	7.05	1.29
CF: Happy BG: Sad	6.86	1.45	6.43	1.42	6.00	1.65	7.45	1.24
CF: Happy BG: Neutral	6.82	1.19	6.56	1.09	6.08	1.38	7.32	1.25
CF: Sad BG: Happy	.71	1.31	1.35	1.26	.71	1.13	.64	1.22
CF: Sad BG: Sad	.31	.52	1.08	1.26	.36	.56	.41	.61
CF: Sad BG: Neutral	.64	.85	1.21	1.14	.57	.78	.65	1.03
CF: Neutral BG: Happy	3.20	1.94	3.64	1.57	2.84	1.58	3.43	1.45
CF: Neutral BG: Sad	1.93	1.64	3.09	1.55	2.35	1.64	3.13	1.60
CF: Neutral BG: Neutral	2.27	1.68	3.18	1.50	2.55	1.63	2.89	1.63
Sadness Judgment								
CF: Happy BG: Happy	.84	1.00	1.92	1.27	.77	.89	.84	1.00
CF: Happy BG: Sad	1.34	1.49	2.28	1.50	1.16	1.42	1.0	1.63
CF: Happy BG: Neutral	1.09	1.05	1.98	1.39	1.07	1.07	1.11	1.35
	7.00	1.73	7.07	1.23	5.88	1.72	7.69	1.48
CF: Sad BG: Happy	7.48	1.46	7.40	1.10	6.44	1.43	7.90	1.01
CF: Sad BG: Sad	6.94	1.26	7.26	1.12	6.34	1.52	7.68	1.35
CF: Sad BG: Neutral	2.13	1.61	3.30	1.58	2.67	1.61	4.48	1.48
CF: Neutral BG: Happy	2.98	2.13	3.65	1.77	3.11	2.08	4.66	1.72
CF: Neutral BG: Sad	1.95	1.57	3.21	1.69	2.43	1.65	4.72	1.71
CF: Neutral BG: Neutral								

Three-Way Analyses of Variance.

Happiness judgments. A 3-way ANOVA looked at the effects of Sample (Japan, Korea, Taiwan, U.S) X Central face (happy, neutral, sad) X Background (happy, neutral, sad) for the judgment of happiness. All main effects and interactions were significant ($p < .001$). The significant 3-way interaction ($F = 2.82$, $p =. 001$) suggested that the samples differ in their judgment of the happiness of the central figure as the background figures differ. The Two-way Analyses of Variances below detail these relationships. For the Happiness judgments, Taiwan and Japan showed similar patterns and Korea and the United States showed similar patterns. The *post hoc* analysis using Tukey HSD showed differences between the U.S. and Japan (mean $\Delta = .36$, $p = .013$) and the U.S and Taiwan (mean $\Delta = .61$, $p < .001$) but the U.S. did not differ significantly from the Korean sample. There were significant differences between Korea and Japan (mean $\Delta = .36$, $p = .014$) and Korea and Taiwan (mean $\Delta = .61$, $p < .001$) but no significance was found between Japan and Taiwan.

Sadness Judgments. Parallel to the happiness judgments, a Sample X Central Face X Background face 3-way ANOVA was undertaken for the sadness judgments. All interactions and main effects were significant ($p < .001$). The significant 3-way interaction ($F = 3.04$, $p <. 001$), suggests that samples differ in their sadness judgment of central figure as the background figures differed. Two-way Analyses of Variances below detail these relationships

The results for sadness judgment with Taiwan and Japan showed similar patterns and Korea and the U.S. showed similar patterns. A *post hoc* analysis using Tukey HSD found significant differences between the U.S. and Japan (Δ mean $= .93$, $p < .001$) and the U.S and Taiwan (mean difference $= 1.13$, $p < .001$) but no difference was found with the Korean sample. There were significant differences between Korea and Japan (mean $\Delta = .70$, $p < .001$) and Korea and Taiwan (mean $\Delta = .91$, $p < .001$) but no significance was found between Japan and Taiwan.

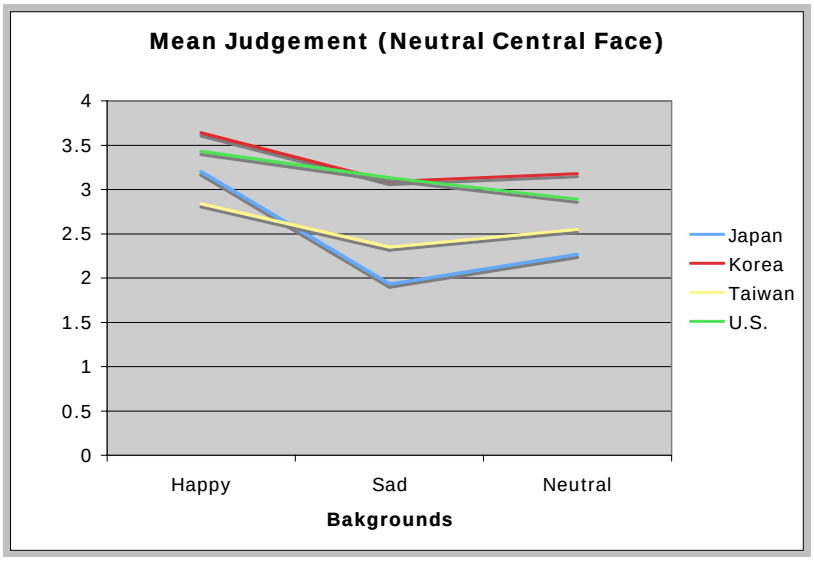
Two-Way Comparisons for Central Faces

Central Face: Happy, Judgment: Happiness. A two-way ANOVA found a main effect for sample in the average judgment of happiness ($F = 22.59$, $p < 0.001$). The U.S sample was higher in their judgment of happiness than were the Japanese, Korean, and Taiwanese samples ($t = 2.58$, $p =. 011$, $t = 5.16$, $p < 0.001$, and $t = 7.45$, $p < 0.001$, respectively). There is no evidence that the background figures made a difference in judgments of 'happiness.' Finally, the interaction between sample and background was not significant showing that samples did not differ in their judgments of the central figure as the background changes.

Central Face: Sadness, Judgment: Happiness. We found the samples to differ in average judgment of happiness ($F = 14.43$, $p < 0.001$). The U.S sample's judgment of happiness was lower than that of the Korean sample ($t= -4.61$, $p < 0.001$) (equal variance not assumed) but was not different from that of the samples from Japan and Taiwan. There was also a main effect of background: the judgment of the happiness differed by background ($F = 20.9$, $p < 0.001$). The interaction between samples and

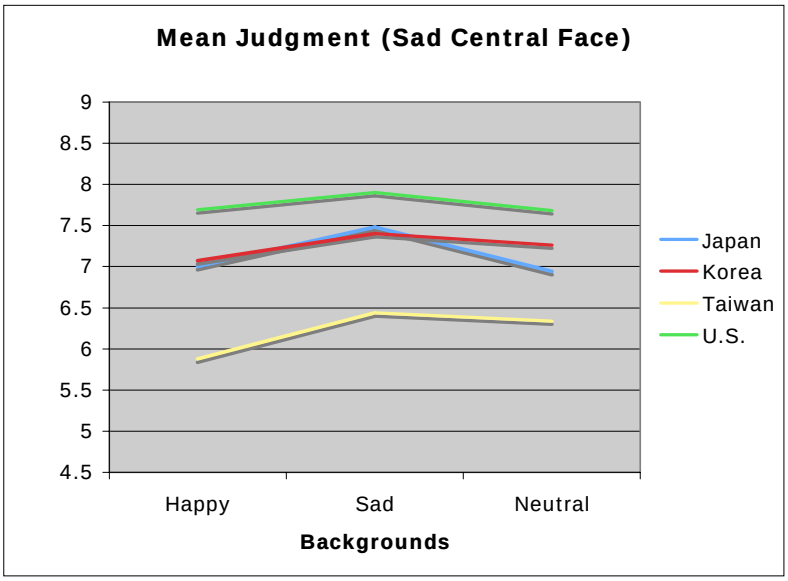
backgrounds was not significant. There is no evidence that samples differed in their judgments as the background changes.

Central Face: Neutral, Judgment: Happiness: This analysis found a significant main effect of sample in the judgment of happiness ($F = 7.71$, $p < 0.001$). The U.S sample's judgment of happiness was higher than that of the Japanese sample ($t = 3.14$, $p = 0.002$) and Taiwanese sample ($t = 2.78$, $p = 0.006$) but was not different from the Korean sample. There was a main effect of background: the judgment of happiness was different among backgrounds ($F = 61.56$, $p < 0.001$). The interaction between samples and backgrounds was significant ($F = 6.80$, $p < 0.001$). The samples differed in their judgments as the background changed. The relationship is depicted in the figure below.

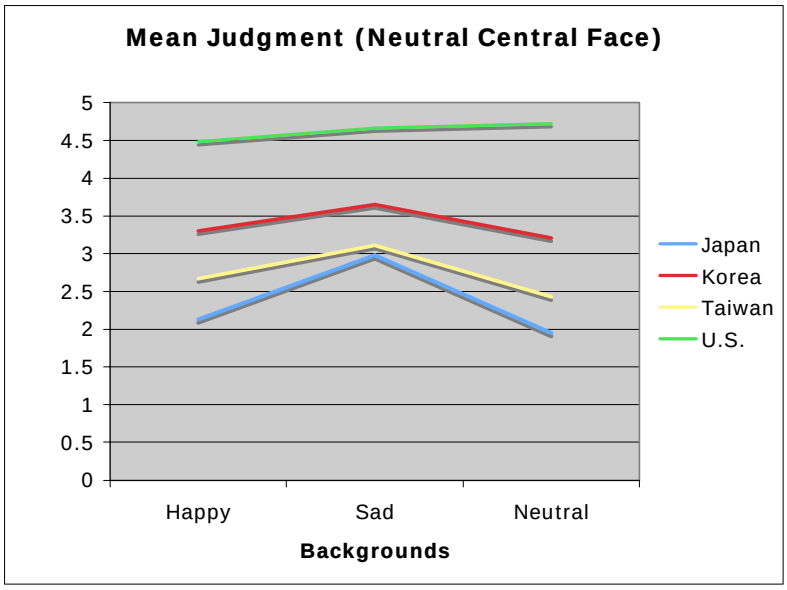


Central Face: Happy, Judgment: Sadness: The samples differed in their average judgment of sadness ($F = 22.97$, $p < 0.001$). The U.S sample was lower in their judgment of sadness than the Korean sample ($t = -6.33$, $p < 0.001$) but not different from the Japanese and Taiwanese samples. There was a main effect of background: the judgment of sadness differed by background ($F = 15.0$, $p < 0.001$). The interaction between samples and backgrounds was not significant.

Central Face: Sadness, Judgment: Sadness: The ANOVA showed that the samples differed in their judgment of sadness ($F = 28.44$, $p < 0.001$). The U.S sample was higher in judgment of sadness than the Japanese, Korean, and Taiwanese samples ($t = 3.62$, $p < 0.001$, $t = 3.36$, $p = 0.001$, and $t = 8.50$, $p < 0.001$ respectively, equal variance not assumed. There was a main effect of background: the judgment of the sadness of the central face differed by background ($F = 20.73$, $p < 0.001$). The interaction between samples and backgrounds was significant ($F = 2.58$, $p = 0.018$). The samples differed in their judgments of the central figure as the background change. The relationship is depicted in the figure below.



Central Face: Neutral, Judgment: Sadness: There was a sample effect in judgment of sadness ($F = 39.91$, $p < 0.001$). The U.S sample was higher in judgment of sadness than were the Japanese, Korean, and Taiwanese samples ($t = 10.35$, $p < 0.001$, $t = 5.67$, $p < 0.001$, and $t = 8.53$, $p < 0.001$, respectively; equal variance not assumed). There was a main effect of background, ($F = 31.06$, $p < 0.001$). Finally, the interaction between samples and backgrounds was significant ($F = 5.46$, $p < 0.001$). The samples differed in their judgments as the background changes. The relationship is depicted in the figure below.



<u>Judgment 'Shakiness'</u>

Are the samples different in the vulnerability of their central figure judgments a function of the emotions of the background figures? The significant interactions for Sample X Background for both happiness and sadness judgments suggests that East Asians' judgment are more vulnerable to background or 'shaky' than are the judgments of the U.S. The U.S. participants, assumed to be analytic thinkers, were expected to show more stable response patterns – their responses would be less influenced by background when the background differed from the central figure.

In order to quantify the overall impact of background on judgment, a measure of judgment shakiness was calculated. The score of 'shakiness' was computed by measuring the absolute value of the variance of their judgment individually. For example, VARHH = (mean Happy Central – HH)^2 +(mean Happy Central – HN)^2+ (mean Happy Central – HS)^2 . One way ANOVA showed significant differences on 'shakiness' scores when Central figure is Neutral for only happiness judgment (VARHN: F = 6.62, p < 0.001) and almost approaching significance for sadness judgment (VARSN: F = 2.60, p = 0.052). In both cases, the Japanese sample showed the most 'shakiness'. They were most sensitive to the background figures.

Shakiness

	Japan		Korea		Taiwan		United States		
Variables	N = 94		N = 92		N = 98		N = 94		
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	F
VARHH	1.70	2.17	1.33	2.78	1.59	3.13	1.32	2.40	.49
VARHN	3.02	4.34	1.69	3.10	1.35	2.74	1.21	1.79	6.62**
VARHS	1.35	5.06	.96	1.90	.75	2.15	1.01	3.62	.50
VARSH	1.65	3.31	1.44	2.17	1.64	4.11	1.83	5.94	.14
VARSN	3.01	3.99	1.79	2.15	2.17	4.36	1.73	3.19	2.60+
VARSS	2.39	4.53	1.12	2.23	1.45	2.39	1.45	4.32	2.25

** p < .001, + p =. 052

Conclusions and Concerns.

The measure appears to reflect social context sensitivity differences. In particular the 3-way ANOVA provides support for the overall role of background figures. The 2-way ANOVAs and the "Shakiness' measure identify the specifics of this impact.

The Japanese team may be able to provide a better review of the data and the analysis as well as interpretation.